Kyle_Casbon_lab3_cp1

April 22, 2025

1 Lab 3

[6]:
```
import io
import os
import glob
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
import scipy as sp
import math
import pandas as pd
```

1.1 Checkpoint 1

1.1.1 1.

Comment on the classification scheme. Did you encounter any ambiguous objects? Show an example (a screenshot is fine) of a galaxy you encountered with clear spiral structure, and one of a galaxy that is morphologically smooth.

I had a very hard time determining whether or not some galaxies had features, especially considering the lower resolution of some of the images. It was also hard to tell what were artifacts and what were real features or points of light.

1.1.2 2.

https://arxiv.org/abs/1308.3496

- What cuts were applied to the SDSS DR7 imaging survey to construct the Galaxy Zoo 2 (GZ2) sample? How many galaxies are in GZ2?
- Describe the galaxy images in GZ2. What are the angular dimensions of the images? Are they the same for all galaxies? Roughly what physical scale in kpc does this correspond to? (You may have to calculate this).
- Describe the classification scheme in GZ2. How many questions are asked for each galaxy?
- How did the classification scheme change from GZ1 to GZ2?
- How does GZ2 decide which galaxies to show a user and in what order?
- How many unique classifications did a typical galaxy get (at the time this paper was written)?

- What steps were taken to deal with unreliable classifications?

- What is the primary reason for apparent redshift evolution (i.e. changes in mean morphological type as a function of redshift) within the GZ2 sample?

- Briefly describe (in words) how GZ2 attempts to correct for classification bias. What is meant by classification bias?

- Why are higher-redshift galaxies in GZ2 more likely to be classified as mergers?

- GZ2 classifies each galaxy using boolean yes/know questions for a finite set of possible classifications. Yet, the reported classifications are all floats (typically ranging from 0 to 1). Why is this?

- Summarize the most significant differences between GZ2 classifications and similar classifications from other morphological catalogs.

1.1.3 3.

Why might astronomers care about morphological classification of galaxies? Check out https://arxiv.org/abs/1403.2783. Give some examples of how morphological classification informs our understanding of the evolutionary history of a galaxy.

1.1.4 4.

Why might astronomers want to systematically search for merging galaxies? Check out https://arxiv.org/abs/1108.2508 and https://arxiv.org/abs/1502.01339. How have merging galaxies been identified in the literature in the past?

1.1.5 5.

How many images are in the training set? What is their dimension?

There are a total of 61,578 images with dimension (424, 424, 3).

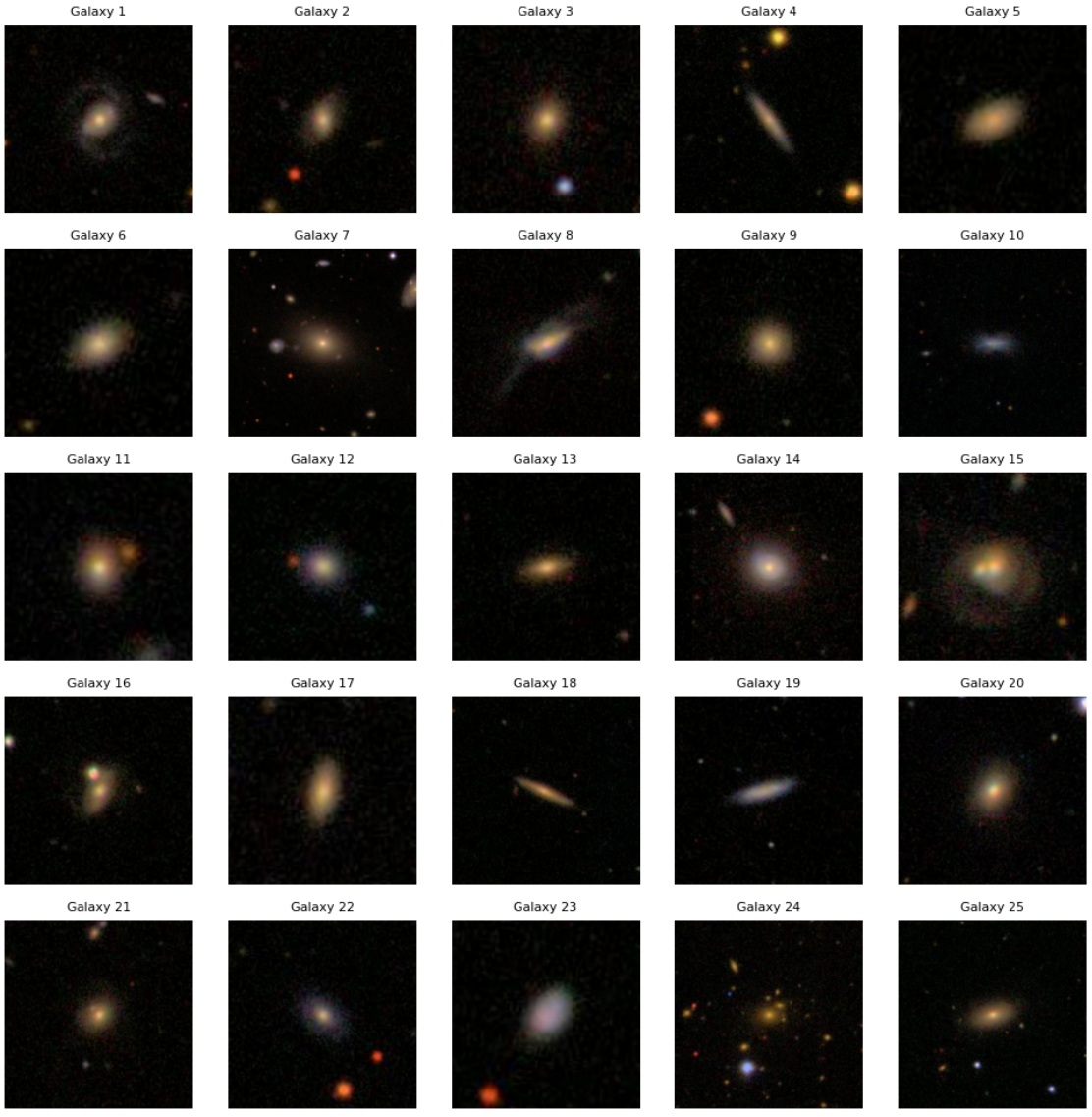
[2]:
```
imgs = ['training_images/100813.jpg', 'training_images/116724.jpg',␣
↪'training_images/123204.jpg', 'training_images/151686.jpg', 'training_images/
↪180544.jpg',
       'training_images/184050.jpg', 'training_images/191850.jpg',␣
↪'training_images/194519.jpg', 'training_images/266051.jpg', 'training_images/
↪285453.jpg',
       'training_images/303434.jpg', 'training_images/310110.jpg',␣
↪'training_images/321865.jpg', 'training_images/342681.jpg', 'training_images/
↪360554.jpg',
       'training_images/375777.jpg', 'training_images/400602.jpg',␣
↪'training_images/414449.jpg', 'training_images/430702.jpg', 'training_images/
↪432101.jpg',
       'training_images/660614.jpg', 'training_images/680092.jpg',␣
↪'training_images/716033.jpg', 'training_images/810107.jpg', 'training_images/
↪999680.jpg',]
```

```
plt.figure(figsize=(10, 10))

for i, img_path in enumerate(imgs):
    img = mpimg.imread(img_path)

    plt.subplot(5, 5, i + 1)
    plt.imshow(img)
    plt.axis('off')
    plt.title(f"Galaxy {i+1}", fontsize=8)

plt.tight_layout()
plt.show()
```



1.1.6 6.

[33]:
```
cl = pd.read_csv('training_classifications.csv')
cols = cl.columns.tolist()
responses = [
    "smooth",
    "features or disk",
    "star or artifact",
    "yes (Could this be a disk viewed edge-on?)",
    "no (Could this be a disk viewed edge-on?)",
    "yes (Is there a sign of a bar feature through the centre of the galaxy?)",
    "no (Is there a sign of a bar feature through the centre of the galaxy?)",
    "yes (Is there any sign of a spiral arm pattern?)",
    "no (Is there any sign of a spiral arm pattern?)",
    "no bulge",
    "just noticeable",
    "obvious",
    "dominant",
    "yes (Is there anything odd?)",
    "no (Is there anything odd?)",
    "completely round",
    "in between",
    "cigar-shaped",
    "ring",
    "lens or arc",
    "disturbed",
    "irregular",
    "other",
    "merger",
    "dust lane",
    "rounded",
    "boxy",
    "no bulge",
    "tight",
    "medium",
    "loose",
    "1",
    "2",
    "3",
    "4",
    "more than four",
    "can't tell"
]

num_plots = 37
cols_per_row = 3
rows = math.ceil(num_plots / cols_per_row)
```

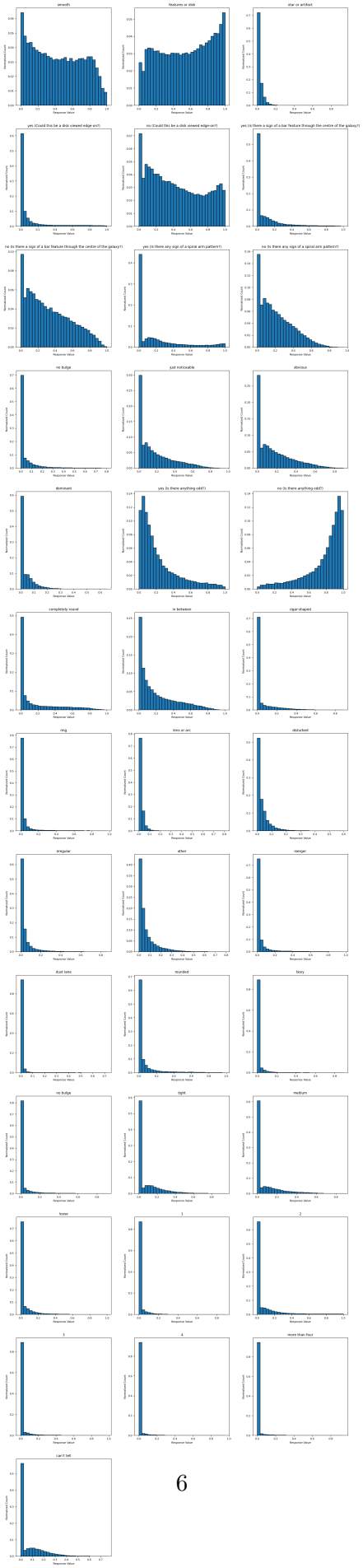
```
fig, axes = plt.subplots(rows, cols_per_row, figsize=(15, 5 * rows))
axes = axes.flatten()

for i in range(num_plots):
    col_data = cl[cols[i + 1]].dropna()  # assuming first column is not one of␣
 ↪the response columns
    counts, bins = np.histogram(col_data, bins=30)
    probabilities = counts / counts.sum()

    ax = axes[i]
    ax.bar(bins[:-1], probabilities, width=np.diff(bins), edgecolor='black',␣
 ↪align='edge')
    ax.set_title(responses[i], fontsize=10)
    ax.set_xlabel("Response Value", fontsize=8)
    ax.set_ylabel("Normalized Count", fontsize=8)
    ax.tick_params(labelsize=8)

for j in range(num_plots, len(axes)):
    axes[j].axis('off')

plt.tight_layout(pad=3.0)
plt.show()
```


smooth
features or disk
star or artifact
yes (Could this be a disk viewed edge-on?)
no (Could this be a disk viewed edge-on?)
yes (Is there a sign of a bar feature through the centre of the galaxy?)
no (Is there a sign of a bar feature through the centre of the galaxy?)
yes (Is there any sign of a spiral arm pattern?)
no (Is there any sign of a spiral arm pattern?)
no bulge
just noticeable
obvious
dominant
yes (Is there anything odd?)
no (Is there anything odd?)
completely round
in between
cigar-shaped
ring
lens or arc
disturbed
irregular
other
merger
dust lane
rounded
boxy
no bulge
tight
medium
loose
1
2
3
4
more than four
can't tell
Response Value
Normalized Count

[34]:
```
folder_path = 'training_images'
jpg_files = [f for f in os.listdir(folder_path) if f.lower().endswith('.jpg')]
print(f"Found {len(jpg_files)} .jpg files in the '{folder_path}' folder.")
```

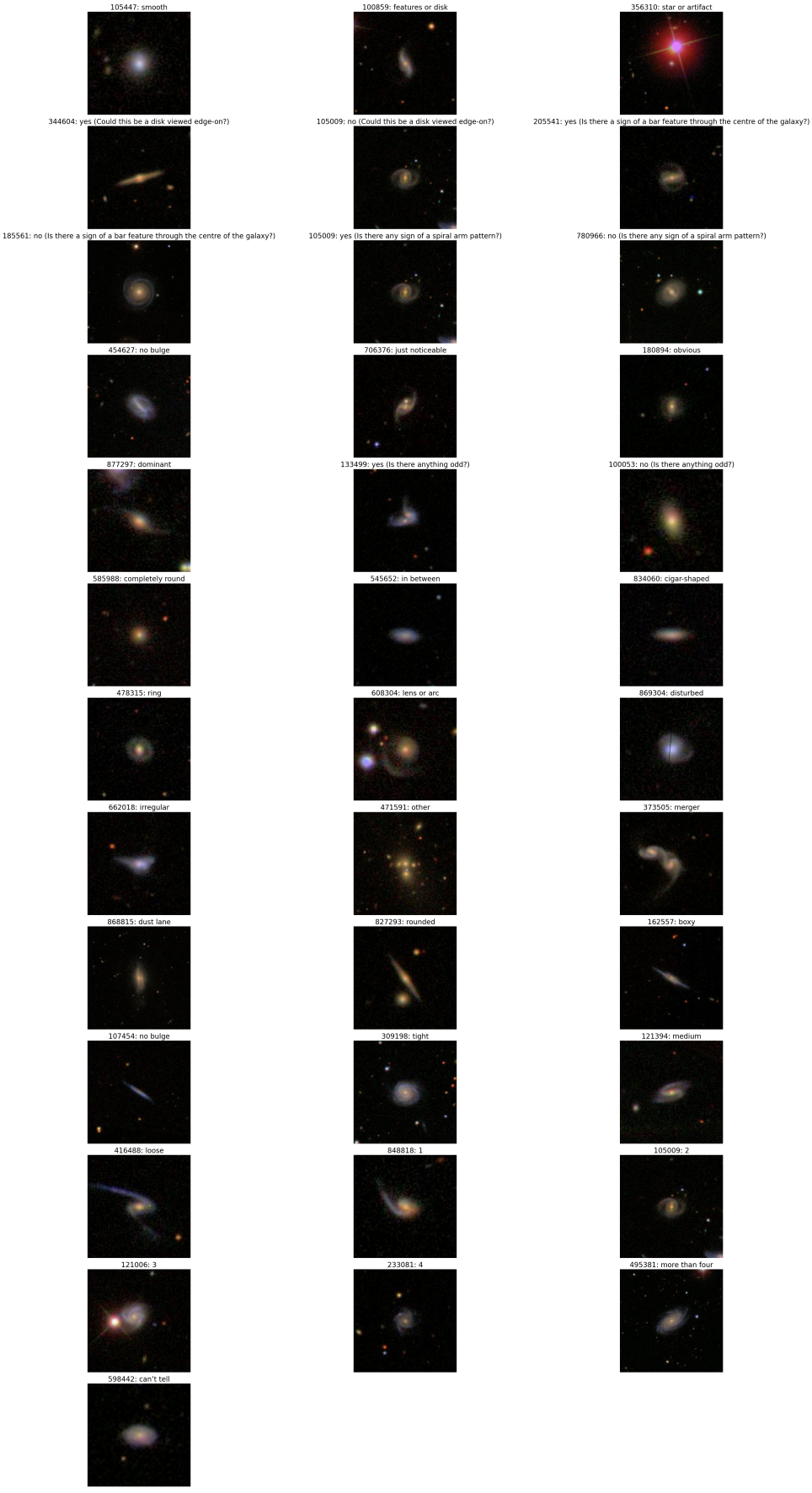
```
Found 75682 .jpg files in the 'training_images' folder.
```

1.1.7 7.

[35]:
```
max_vals = []
max_ids = []
image_paths = []
for i in range(1, 38):
    res = f"{responses[i-1]}"
    col = cols[i]
    max_row = cl.loc[cl[col].idxmax()]
    max_vals.append(max_row[col])
    max_id = int(max_row['GalaxyID'])
    max_ids.append(max_id)
    image_paths.append(f"training_images/{max_id}.jpg")
```

[36]:
```
fig, axes = plt.subplots(14, 3, figsize=(32, 64))
axes = axes.flatten()

for i, ax in enumerate(axes):
    if i < len(image_paths):
        img = mpimg.imread(image_paths[i])
        ax.imshow(img)
        ax.axis('off')
        ax.set_title(f"{max_ids[i]}: {responses[i]}", fontsize=20)
    else:
        ax.axis('off')
plt.tight_layout()
plt.show()
```


105447: smooth
100859: features or disk
356310: star or artifact
344604: yes (Could this be a disk viewed edge-on?)
105009: no (Could this be a disk viewed edge-on?)
205541: yes (Is there a sign of a bar feature through the centre of the galaxy?)
185561: no (Is there a sign of a bar feature through the centre of the galaxy?)
105009: yes (Is there any sign of a spiral arm pattern?)
780966: no (Is there any sign of a spiral arm pattern?)
454627: no bulge
706376: just noticeable
180894: obvious
877297: dominant
133499: yes (Is there anything odd?)
100053: no (Is there anything odd?)
585988: completely round
545652: in between
834060: cigar-shaped
478315: ring
608304: lens or arc
869304: disturbed
662018: irregular
471591: other
373505: merger
868815: dust lane
827293: rounded
162557: boxy
107454: no bulge
309198: tight
121394: medium
416488: loose
848818: 1
105009: 2
121006: 3
233081: 4
495381: more than four
598442: can't tell